Linearity in the loudness envelope of the *messa di voce*

Densil Cabrera (1), Manuj Yadav (1) and Dianna T. Kenny (2)

(1) Faculty of Architecture, Design and Planning, The University of Sydney, NSW 2006, Australia
(2) Faculty of Arts, The University of Sydney, NSW 2006, Australia

ABSTRACT

The *messa di voce* (MDV) is a vocal exercise used by singers, consisting of a crescendo and decrescendo on a single sustained note. In this study we analysed recorded MDVs sung by tertiary singing students to examine the extent to which MDVs have a linear crescendo and linear decrescendo. The MDVs studied were recorded over a 3-year period, as a cohort of classical singing students progressed through their tertiary education and training. Previous studies of MDV envelopes have examined the envelopes in decibels, but in the present study we use envelopes derived from a dynamic loudness model. We did not find an overall tendency for increased linearity as students mature.

INTRODUCTION

Classical singing at a professional level typically requires many years of training, so that the singer has the ability to project their voice appropriately, and control their voice precisely. There are many singing exercises that can be used to help a student refine their voice control, and the *messa di voce* (Italian for 'placing the voice', and abbreviated here to MDV) is one such exercise. The MDV is very simple in concept, consisting of a crescendo and decrescendo on a single sustained note (while simple in concept, the MDV is not simple for a singer to execute). The simplicity of the MDV makes it an interesting candidate for acoustic analysis of voice production, exemplified by several previous studies (e.g., Titze 1992, Titze *et al.* 1999, Bretos and Sundberg 2003, Collyer *et al.* 2007, Collyer *et al.* 2009, Mitchell and Kenny 2010). Such analysis can examine both coarse and fine features of MDVs – such as envelope symmetry, the linearity of the growth and decay functions, the voice spectrum (e.g., how it changes with sound pressure level and pitch), vibrato, and so on. In this study we analyse a set of recorded MDVs to examine the extent to which their loudness envelopes exhibit linear growth and decay functions.

Two previous studies have focussed on the envelope structure of MDVs. Titze *et al.* (1999) studied MDVs from six singers in relation to physiological features (such as lung volume), finding that greater temporal symmetry in the MDV sound pressure level was evident in participants who had a smaller dynamic range. They suggest that this could be because the high dynamic range participants expend more of their lung volume, giving them less control. The high dynamic range MDVs tended to be characterised by a delayed rise in sound pressure level, followed by a sudden fall after the peak. Titze *et al.* (1999) also observed that some MDVs have a plateau-like sound pressure level envelope (i.e., the maximum level is sustained for some time), and they speculated that even though the level did not vary much during this period, the loudness of the MDV might still be changing due to the effects of vibrato or spectrum (i.e. changes in the strengths of formants). This observation is relevant to the present study, in which we examine the MDV envelope using a computational loudness model.

The other major study of MDV envelope structure is by Collyer *et al.* (2007), using five singers. Their study had a stronger focus on the shape of the crescendo and decrescendo envelope, that is, the extent to which it is linear. Unlike Titze *et al.*, they did not observe a relationship between dynamic range and linearity. They examined the relationship between sound pressure level and spectral balance (expressed as the ratio of power in the 0-2 kHz band to that in the 2-4 kHz band), finding a linear correspondence.

According to Titze *et al.* (1999), the ideal MDV has a symmetric triangular envelope – and this ideal was taught to the singers involved in the present study. However, one of the difficulties with previous acoustical studies of MDV envelopes is that the envelopes are represented in decibels. The identification of linearity in crescendo and decrescendo is somewhat problematic using the decibel scale, since the scale has no true zero and is not linearly related to loudness. It seems unlikely that singers would aim to perform in relation to the decibel scale – especially as most singers would be unfamiliar with it. The 'plateau' mentioned by Titze *et al.* might be partly due to the compressive effect that a logarithmic scale has on high underlying values. This raises the question, then, of what might we mean by 'linear'. There are many possibilities – for example, we could consider the pressure envelope or the pressure squared envelope, both of which are common ways of representing physical sound quantity without using decibels. Another possibility is to raise the pressure envelope to the power of 0.6, which is the exponent found by Stevens (1955) relating the pressure of mid-frequency pure tones to loudness. A reasonable assumption is that in aiming for a linear crescendo and decrescendo, singers attempt to control the *loudness* of their voice linearly (as suggested by Titze *et al.*). If that is the case, then some type of loudness model should be effective for analysing the MDV envelope, and there are more sophisticated approaches to this than raising the pressure envelope to a power.

Time-varying loudness can be modelled in various ways, and we have previously used the models of Glasberg and Moore (2002) and Chalupper and Fastl (2002), obtaining similar results from the two models (Lee and Cabrera 2010). Such models include the effects of the outer and middle ear transfer functions, auditory filtering in the inner ear, functions relating excitation to specific loudness, temporal integration, and loudness summation. Differences between these two models are examined by Rennies *et al.* (2010). For the sake of succinctness, rather than focusing on detailed issues in loudness modelling, in this paper we present results derived from Glasberg and Moore's (2002) model.

The particulars of how we applied a loudness model to MDV recordings are given in the Method section. However, it is worth considering at this point what a loudness model is likely to achieve. Firstly, because of temporal integration, the loudness envelope will be relatively smooth – and vibrato will have a relatively small effect on envelope fluctuations. In fact, vibrato has the potential to increase the calculated loudness because a frequency-modulated signal can excite a broader range of auditory filters. The strength of formants in the voice, too, may affect the modelled loudness from their position relative to the outer and middle transfer function peaks, and from their potential to change the excitation pattern of the inner ear. As such characteristics of the voice could change during the course of an MDV, a loudness model is likely to yield a significantly different envelope to a pressure envelope.

METHOD

The data analysed for this paper comes from a 3-year longitudinal study in which tertiary classical singing students were recorded singing MDVs once every six months. In a recording session, each student performed the MDVs on each of four pitches. The four pitches were in the form of a root position major triad and upper octave tonic, and from lowest to highest pitch, these are referred to as MDV1 to MDV4. The participant group included four voice types: baritone, tenor, mezzo-soprano and soprano. There were 28 participants, of whom 15 completed the three-year study.

Recordings were made with a Brüel & Kjær type 4128 microphone positioned 7 cm from the corner of the mouth. This microphone position has been used in several singing studies following Cabrera *et al.* (2002). Calibration tones were recorded, so that all the recordings were matched in gain, and sound pressure level could be determined. More details of the recording procedure and participants are given by Ferguson *et al.* (2010).

In order to avoid including the silence before and after each MDV in our analysis, we analysed each one for its pitch strength, using an algorithm called SWIPE' (Camacho and Harris 2008). This provided a robust indicator of voiced sound, which allowed automated truncation of unwanted parts of each sound recording.

A total of 437 MDVs were recorded *and* analysed (as described in this paper) from the study – some technical problems prevented the analysis of certain recordings. We were able to generate complete data sets for ten participants (from the 12 participants who had MDVs recorded in all six semesters), and the focus of our analysis was on these ten.

Calculating loudness is not a simple matter, and in this situation, there is an additional complication: it would be inappropriate to calculate loudness at the measurement microphone position because it is inconceivable that someone would listen to singing at that position. Therefore, we attenuated the signal as follows. The maximum sound pressure level of each MDV was derived, and the power average of these maxima was determined. Attenuation was then applied to every MDV recording such that the average of the maxima was equal to 80 dB (the same attenuation was applied to all recordings prior to loudness analysis, so that their relative levels were preserved). The choice of this sound pressure level is based on the study of Cabrera *et al.* (in press 2011), which showed that the sound pressure level (L_{eq}) measured from a solo singer in a recital hall was typically 75-85 dB (singing a song with a wide dynamic range, over a duration of 40-60 s, with the performer on stage and microphone in the audience area). Hence, we are examining the extent to which the loudness of the MDV in a plausible listening situation follows the ideal triangular envelope.

The attenuated recordings were analysed using Glasberg and Moore's time-varying loudness model (2002), as implemented in PsySound3 (Cabrera *et al.* 2008), which is a Matlab-based sound analysis environment. We used the outer and middle ear transfer functions given in Glasberg and Moore's paper, as it is difficult to define a particular outer ear transfer function for our situation. Furthermore, in our previous work with loudness modelling of room impulse response slopes, we found that the spectral weighting of the signal did not have a significant effect on loudness decay (and presumably growth) functions (Lee and Cabrera 2010), and this is also likely to be the case with MDV analysis – since we are, likewise, analysing the slope of the envelope rather than the envelope in absolute terms. From the loudness model, we used short term loudness (Glasberg and Moore 2002), which models the time-varying loudness that could be mentally tracked by a listener (as opposed to the long term loudness, which provides an indication of the overall loudness).

In order to derive a measure of how triangular each MDV's loudness envelope was, we performed a curve-fit against a function with a linear increase, joined to a linear decrease (Equation 1). In this equation: $N(t)$ is loudness as a function of time; t is time; a controls the steepness of both sides of the envelope; b is equal to the time (t value) at which the function peaks; c is used to introduce the possibility of asymmetry in the envelope slope by tilting the triangle; and ab shifts the function into the positive (because loudness values are positive), ensuring that the function has a value of 0 at t=0. Curve fitting was performed in Matlab (using the Curve Fitting Toolbox).

$$N(t) = a|t-b|+ct-ab. \tag{1}$$

After a period of experimentation, we decided to fix the value of b to the time at which maximum loudness was reached, leaving just two unknowns in the equation (a and c). We used this approach instead of modelling two linear regressions (one each for crescendo and decrescendo) because if separate regressions are modelled they do not necessarily meet at the peak (we found some instances of substantial steps between pairs of linear regressions when we tried this, for example, when there is a sudden drop in loudness just after the maximum). There are further alternatives to modelling too, depending on the constraints that one wishes to apply to the model (for example, whether the peak time should be the actual peak or one derived from function fitting, and whether the start and end of the fitted function should equal zero sones or be unconstrained).

Goodness of fit can be assessed in many ways, and in this context we simply used the root mean square error divided by the mean value of the data (this is very similar to the coefficient of variance). The reason for dividing by mean is that MDVs that have a higher maximum sound pressure level tend to have higher deviation – and this could be expected if loudness is a positive ratio scale. Doing this results in similar errors for MDV1 to MDV4, whereas if rms error is used without dividing by mean, the error increases with the pitch of the MDV.

RESULTS

Two examples of results of individual participants are given (Figures 1 and 2). Figure 1 shows the loudness envelopes (MDV2) for each of the six semesters, for a participant whose MDV shows a wide variety of envelope shapes. The MDV in the final semester is closest to a linear crescendo and decrescendo, but these six MDVs (considered alone) do not show a convincing trend towards increased linearity. The respective error value and fitted function are shown on each chart.

Figure 2 give an example of a participant whose first MDV recording exhibits a highly linear loudness envelope, with all subsequent MDVs straying further from the ideal triangular function (although in the fifth semester the MDV is almost as linear as first semester). This example lends support to the notion that the triangular loudness envelope could be an ideal, but it shows a participant whose result was so well matched to the ideal at the outset that there was no scope for improvement.

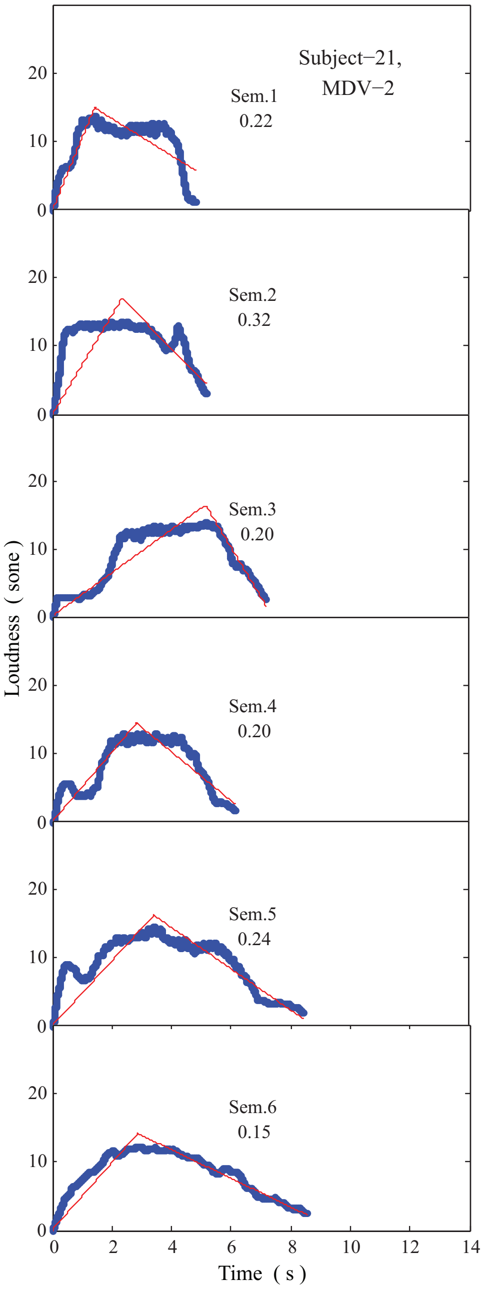

Figure 1. Example of a set of MDVs recorded over 6 semesters. The thick blue trace represents the original data, as modeled by the loudness algorithm, and the red line is the fitted function. Error values are shown in each of the six charts. This chart is for participant 21, on MDV2.

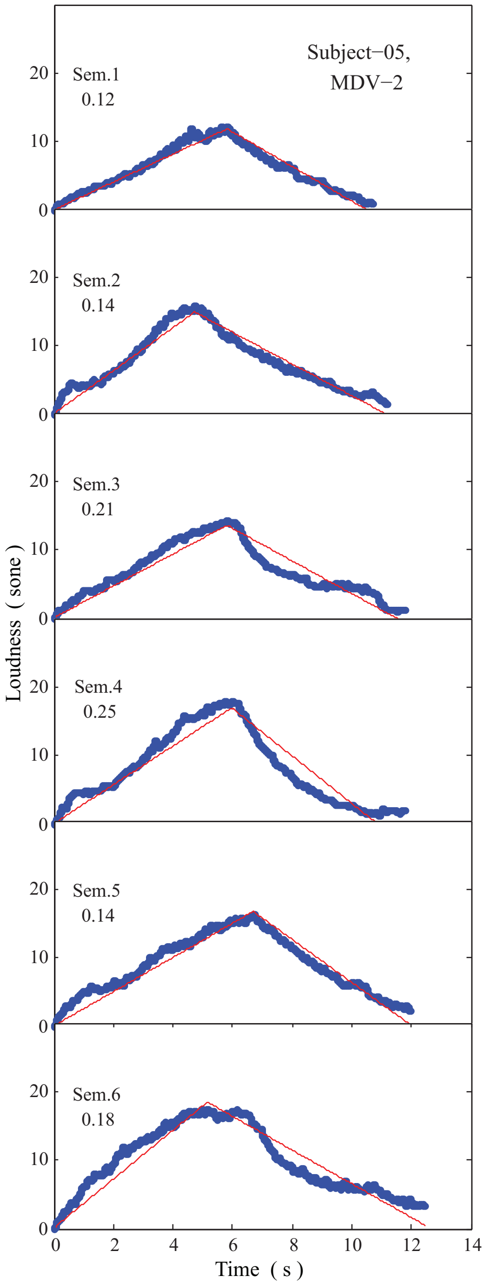

Figure 2. Example of a set of MDVs recorded over 6 semesters. The thick blue trace represents the original data, as modeled by the loudness algorithm, and the red line is the fitted function. Error values are shown in each of the six charts. This chart is for participant 5, on MDV2.

Overall, the analysis does not support the hypothesis that the loudness envelope becomes more linear over three years of training. Triangular function fitting was performed on every MDV loudness envelope, from which the root mean square error divided by the data mean was determined. If linearity increases over the period of training, we should see this error reduce over time, but instead there is no clear trend in the mean of Figure 3 (shown by the heavy red line). Error results for the participants who were analysed in all six semesters are also shown individually in Figure 3, with no general trend evident.

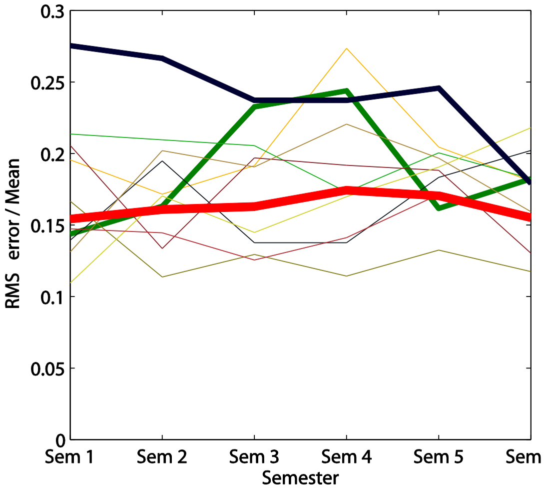

Figure 3. Error from curve-fitting the loudness envelopes of those participants who recorded a complete set of MDVs over the six semesters. Each line represents the results of a single participant (with errors from MDV1 to MDV4 averaged). Participant 21 (see Fig 1) is shown given by the heavy blue line, Participant 5 (see Fig. 2) is given by the heavy green line, and the mean error across participants is given by the heavy red line.

As indicated by Figure 3 the highly triangular MDV envelopes exemplified by Participant 5 (Fig. 2) were not an isolated exception: similarly good-fitting results were found in MDVs from other participants. For guidance in interpreting the error values in Figure 3, consider that Participant 5's Semester 1 and 5 errors were 0.12 and 0.14 respectively. On the other hand the irregular MDV envelopes of Participant 21 (Fig. 1) were exceptionally so. The data of Figure 3 includes all four MDVs for Participant 21, and it can be seen this participant's MDVs become closer to the norm over the period of study (although the mean difference is not significant).

DISCUSSION

Previously we have analysed many other features of the recorded MDVs, to see if there are tendencies over the six semesters. We examined the sound pressure level distribution (especially maximum and median of each MDV), spectral energy distribution (including short-term energy ratio statistics, as defined by Ferguson *et al.* 2010), energy (the product of pressure squared and MDV duration), and vibrato parameters (S. Ferguson, D. Kenny, H. Mitchell, M. Ryan and D. Cabrera – manuscript in preparation 2011). For most of the parameters, there are discernible patterns over the course of the six semesters, some statistically significant. Hence, the lack of a discernible tendency in the present study does not mean that the MDV set of a student cohort has no evolution over six semesters of voice training. It does indicate that the linearity of the loudness envelope in crescendo and decrescendo may not systematically change, at least using the assumptions of the present study.

Our curve fitting model did not penalise envelope asymmetry – i.e., the peak of the envelope could be near the start or the end of the MDV, and if the crescendo to the peak and decrescendo from the peak were linear, then a low error value would be returned. However, such asymmetry is clearly undesired if the aim of the MDV is as expressed by Titze *et al.* (1999). Of course, there is little need to employ detailed loudness modelling to examine whether the peak is in the middle of the MDV period, and we chose not to confound symmetry with linearity. Nevertheless, a further analysis could be done with a triangular envelope peaking in the middle of the MDV period, or at the centroid of the MDV envelope, which might yield more positive results.

We examined several other approaches to curve-fitting a linear crescendo and decrescendo (*without* centering the peak), but none of the other approaches yielded more promising results. Hence, even though there is some scope for improving the current fitting function (for example, by anchoring the final datapoint on zero sones – which would provide a better error model for the first MDV of Figure 1), it seems unlikely that a different conclusion would be drawn.

As part of ongoing work, we are conducting a similar analysis of the MDV set to that described in this paper, but modeling the loudness that the singer hears of her/his own voice. For that, the transfer function from measurement microphone to the auditory system needs to be defined, accounting for both airborne and structure-borne (bone conducted) components, as well as changes in auditory sensitivity during singing.

CONCLUSION

This study finds that increased linearity in the loudness envelope of MDVs does not evolve over three years of tertiary singing training, considering ten singers as a group. This suggests that, in general, linearity in loudness control over the MDV is not learnt during tertiary level training (which is not to say that it is not learnt prior to tertiary training, as evidenced by the data in Figure 2). Nevertheless, the concept of analysing MDVs using a computational loudness model has some potential for better underdstanding of the singing voice.

ACKNOWLEDGEMENTS

Support for this research was provided by an Australian Research Council's Discovery Projects Grant to Dianna Kenny, Helen Mitchell, Densil Cabrera and Michael Halliwell (DP0558186). The MDV recordings were made by Helen Mitchell, and advice on analysis was provided by Sam Ferguson.

REFERENCES

Cabrera, D, Davis, PJ, Barnes, J, Jacobs, M and Bell, D 2002, 'Recording the operatic voice for acoustic analysis', *Acoustics Australia*, vol. 30, no. 3, pp. 103-108.

Cabrera, D, Ferguson, S and Schubert, E 2008, 'PsySound3: an integrated environment for the analysis of sound recordings', in *Acoustics 2008: proceedings of the Australian Acoustical Society conference*, Geelong, Australia

Cabrera, D, Davis, PJ and Connolly A (in press 2011) 'Long-term horizontal vocal directivity of opera singers: effects of singing projection and acoustic environment', *Journal of Voice*.

Camacho, A and Harris, JG (2008) ‘A sawtooth waveform inspired pitch estimator for speech and music’, *Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1638-1652.

Challuper, J and Fastl, H 2002, ‘Dynamic loudness model (DLM) for normal and hearing-impaired listeners’, *Acta Acustica united with Acustica*, vol. 88, pp. 378-386.

Collyer, S, Davis, PJ, Thorpe, CW & Callaghan, J 2007 ‘Sound pressure level and spectral balance linearity and symmetry in the messa di voce of female classical singers’, *Journal of the Acoustical Society of America*, vol. 121, pp. 1728–1736.

Collyer, S, Davis, PJ, Thorpe, CW & Callaghan, J 2009, ‘Fundamental frequency influences the relationship between sound pressure level and spectral balance in female classically trained singers’, *Journal of the Acoustical Society of America*, vol. 126, pp. 396-406.

Ferguson, S, Kenny, D and Cabrera, D 2010 ‘Effects of training on time-varying spectral energy and sound pressure level in nine male classical singers’, *Journal of Voice*, vol. 24, no. 1, pp. 39-46.

Glasberg, BR and Moore, BCJ 2002, ‘A model of loudness applicable to time-varying sounds’, *Journal of the Audio Engineering Society*, vol. 50, no.5, pp.331-342.

Lee, D and Cabrera, D 2010, ‘Effect of listening level and background noise on the subjective decay rate of room impulse responses: Using time-varying loudness to model reverberance’, *Applied Acoustics*, vol. 71, pp. 801-11.

Mitchell, HF and Kenny, DT 2010, ‘Change in vibrato rate and extent during tertiary training in classical singing students’, *Journal of Voice*, vol. 24, no. 4, pp. 427-434.

Rennies, J, Verhey, JL and Fastl, H 2010, ‘Comparison of loudness models for time-varying sounds’, *Acta Acustica united with Acustica*, vol. 92, pp. 383-396.

Stevens, SS 1955, ‘The measurement of loudness’, *Journal of the Acoustical Society of America*, vol. 27, pp. 815-829.

Bretos, J and Sundberg, J 2003, ‘Measurements of vibrato parameters in long sustained crescendo notes as sung by ten sopranos’, *Journal of Voice*, vol. 17, no. 3, pp.343-352.

Titze, IR 1992 ‘*Messa di voce*’, *Journal of Singing*, vol. 48, pp. 24-25.

Titze, IR, Long, R, Shirley, GI, Stathopoulos, E, Ramig, LO, Carroll, LM and Riley, WD 1999, ‘*Messa di voce*: An investigation of the symmetry of crescendo and decrescendo in a singing exercise’, *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2933-2940.